B-108E
C.2.

Statistics Canada Statistique Canada

Ottawa

DESIGN OF THE 1982 POULTRY & EGG SURVEY

K. Smith-Doiron
J-L Tambay
IASMD

December, 1982

Table of Contents

1. Introduction

2. Sampling Frame

3. Sample Design

4. Overlap with N.F.S.

5. Estimation Procedures

Appendix I - Design Specifications

Appendix II - Formulae for Estimation

Design of the 1982 Poultry & Egg Survey

1. Introduction

The Poultry and Egg Survey was formerly a monthly mail-out survey with low response rates. The re-designed survey (since Oct. 82) is a quarterly mail-out survey with telephone follow-up of non-respondents (except for farms with less than 18 layers.) There are two types of farms in the poultry industry: registered and non-registered. Registered flocks are those for which the number of layers are equal to or over the limit designated in each province: 500 in Nfld, NS, Ont, Man and BC; 300 in PEI, Sask and Alta; 250 in Quebec and 200 in NB. There will be no survey of registered flocks in the Atlantic provinces because of the small number of such producers there for whom timely information is available from the provincial boards.

There will be a 25% rotation of the sample every October except in the complete enumeration strata. The sample design and sample size should lead to an expected design C.V. of 5% for estimates of the numbers of layers at the province level, separately for registered and non-registered flocks. The CV's for other variables may not be in the same range.

2. Sampling Frame

The 1981 Census of Agriculture was used as a base. Indian Reserves, institutions and farms with zero layers were removed from the frame. In the Atlantic provinces registered flocks were also removed.

3. Sample Design

Non-Registered Flocks

In each province, the non-registered flocks were stratified using the same classifications used in Census publications. The sample size was allocated using optimum allocation. In one case (stratum 4 in Newfoundland), the optimum allocation demanded complete enumeration. The farms in that stratum are called specified farms.

Registered Flocks

The registered flocks were broken down into three strata within each province. The stratum boundaries were determined in an optimum manner for optimum allocation. Again this leads to some strata being completely enumerated (i.e. stratum 6 in Quebec, stratum 7 in Ontario and stratum 7 in Saskatchewan).

The sample size (except in complete enumeration strata) is a multiple of four to allow 25% rotation. There will be no rotation in the complete enumeration strata. A simple random sample of farms was selected in each stratum. The stratum definitions and population and sample sizes are given in Appendix I Table I.

4. Overlap with National Farm Survey

The Poultry sample will be matched to the 1983 NFS specified and list sample. Any overlap poultry farms will be identified and their July questionnaires will be placed in the NFS assignment packages. These questionnaires will be completed by an NFS interviewer at the same time as NFS. In subsequent years, the common NFS area farms can also be identified and overlap poultry questionnaires can be inserted in the assignment packages.

5. Estimation Procedures

Estimation procedures and formulae for calculating the level estimates and their estimates of variation are given in Appendix II.

The estimators account for both the sample design of the survey and anticipated nonresponse to the survey. The estimators associated with the sample design used (i.e. a simple random sample within each stratum) were incorporated with a model for nonresponse to develop the survey estimators. A brief description of the nonresponse model follows.

The model assumes that each farm, once contacted for the survey, will report a particular item or group of items on the questionnaire with a certain probability. That probability may not be the same for all questionnaire items or for all farms in the survey, but it will not change over the duration of the survey for the same farm. The nonresponse model also imposes some restrictions on the values that the probability of response will take.

The first restriction is that out-of-business farms will always report all items on the questionnaire if contacted, the reported values being zeroes. This restriction is based on the assumption that all out-of-business farms in the sample will be correctly identified as such at the data collection stage, hence their probabilities of response are always equal to one. A second restriction is that all the farms in a stratum, excluding out-of-business farms, have the same probability of response to a particular item or group of items on the questionnaire.

The underlying assumption is that farms within a stratum share certain characteristics that will make them behave in a similar way if in the sample. Finally, the reporting status of a farm with respect to a particular item or group of items will in no way influence the reporting of those items by another farm in the survey. This should hold in most survey situations.

During the development of the survey estimators, the probabilities of response are treated as if they were known constants. In actual case, the values are estimated using respondents to the survey. This is not expected to cause significant changes to the results of the survey.

APPENDIX I

TABLE I

1982 POULTRY & EGG SURVEY SAMPLE DESIGN

PROVINCE	TYPE OF FARM[1]	STRATUM	# OF LAYERS	POPULATION SIZE	SAMPLE SIZE
Newfoundland	NR	1	1-17	58	4
		2	18-77	43	12
		3	78-249	7	4
		4	250-499	6	6
		TOTAL		114	26
Prince Edward Island	NR	1	1-17	256	8
		2	18-77	217	20
		3	78-299	63	24
		TOTAL		536	52
Nova Scotia	NR	1	1-77	512	12
		2	18-77	450	24
		3	78-249	56	12
		4	250-499	18	8
		TOTAL		1,036	56
New Brunswick	NR	1	1-17	326	8
		2	18-77	426	28
		3	78-199	96	16
		TOTAL		848	52
Quebec	NR	1	1-17	2,905	12
		2	18-77	3,461	40
		3	78-249	517	20
		SUB-TOTAL		6,883	72
	R	4	250-11,569	269	32
		5	11,570-52,329	89	28
		6	52,330 and over	9	9
		SUB TOTAL		367	69
		TOTAL		7,250	141
Ontario	NR	1	1-17	6,559	8
		2	18-77	6,767	20
		3	78-249	1,503	12
		4	250-499	652	12
		SUB-TOTAL		15,481	52
	R	5	500-8,254	703	36
		6	8,255-36,658	281	36
		7	36,659 and over	28	28
		SUB-TOTAL		1,012	100
		TOTAL		16,493	152

[1] R=registered flocks
NR=non-registered flocks

PROVINCE	TYPE OF FARM[1]	STRATUM	# OF LAYERS	POPULATION SIZE	SAMPLE SIZE
Manitoba	NR	1	1-17	1,337	4
		2	18-77	2,445	28
		3	78-249	347	12
		4	250-499	121	8
		SUB-TOTAL		4,250	52
	R	5	500-6,151	163	16
		6	6,152-16,977	88	12
		7	16,978 and over	31	24
		SUB-TOTAL		282	52
		TOTAL		4,532	104
Saskatchewan	NR	1	1-17	4,286	8
		2	18-77	8,507	48
		3	78-299	715	16
		SUB-TOTAL		13,508	72
	R	4	300-3,558	92	12
		5	3,559-16,595	45	16
		6	16,596-114,999	9	4
		7	115,000 and over	1	1
		SUB-TOTAL		147	33
		TOTAL		13,655	105
Alberta	NR	1	1-17	3,575	8
		2	18-77	8,146	40
		3	78-299	1,159	24
		SUB-TOTAL		12,880	72
	R	4	300-3,701	260	20
		5	3,702-13,564	140	28
		6	13,565 and over	28	16
		SUB-TOTAL		428	64
		TOTAL		13,308	136
British Columbia	NR	1	1-17	2,888	12
		2	18-77	3,369	36
		3	78-249	265	8
		4	250-499	69	8
		SUB-TOTAL		6,591	64
	R	5	500-10,212	99	12
		6	10,213-21,862	110	12
		7	21,863 and over	16	8
		SUB-TOTAL		225	32
		TOTAL		6,816	96
Canada	NR+R	TOTAL		64,588	920

[1] R=registered flocks
NR=non-registered flocks

APPENDIX II: Formulae for Estimation

The estimators and their estimated variances are developed in the context of the sample design and the nonresponse model used. Section A of this appendix describes the notation used, Section B the assumptions in the model and Section C the theoretical results that lead to the derivation of the Formulae IV to VII.

In order to simplify the notation, the sample will be treated as if it came from only one stratum until otherwise indicated.

A. Notation:

y_i - the value of variable y for the i^{th} unit in the population (i = 1, 2, ..., N)

x_i - the value of variable x for the i^{th} unit in the population (i = 1, 2, ..., N)

N - the population size

n - the sample size

a_i - random variable used to indicate if unit i was in the sample

$$a_i = \begin{cases} 1 & \text{if unit i was in the sample} \\ 0 & \text{otherwise} \end{cases}$$

NOTE: $\sum_{i=1}^{N} a_i = n$

r_{yi} - response random variable used to indicate if sampled unit i responded to the question on variable y,

$$r_{yi} = \begin{cases} 1 & \text{if unit i responded to variable y} \\ 0 & \text{otherwise} \end{cases}$$

r_{xi} - response random variable used to indicate if sampled unit i responded to the question on variable x

$$r_{xi} = \begin{cases} 1 & \text{if unit i responded to variable x} \\ 0 & \text{otherwise} \end{cases}$$

Y - population total for variable y

$$Y = \sum_{i=1}^{N} y_i$$

X - population total for variable x

$$X = \sum_{i=1}^{N} x_i$$

B. Assumptions:

1. The sample is a simple random sample without replacement. This gives the following standard results:

$$Pr(a_i = 1) = \frac{n}{N}$$

$$Pr(a_i a_j = 1) = \frac{n(n-1)}{N(N-1)} \quad \text{for } i \neq j$$

2. Each unit in the population belongs to one of two groups denoted by I and 0, representing farms in business and out-of-business farms, respectively. The values of variables y and x are always zero for units in group 0 and these units always respond to variables y and x. i.e. For $i \varepsilon 0$, $y_i = x_i = 0$ and

$$r_{yi} = r_{xi} = 1$$

3. For farms in Group I given that $a_i = 1$, we have

$$Pr(r_{yi} = 1 \;/\; i \varepsilon I) = p_y,$$

$$Pr(r_{xi} = 1 \;/\; i \varepsilon I) = p_x, \quad \text{and}$$

$$Pr(r_{yi} = 1, \; r_{xi} = 1 \;/\; i \varepsilon I) = p.$$

4. Given a_i and a_j, the response random variables r_{yi} and r_{xi} are independent of the response random variables r_{yj} and r_{xj} for $i \neq j$.

C. Results:

1. $E(r_{yi} y_i / a_i = 1) = p_y y_i$, since $y_i = 0$ when $i \varepsilon 0$.

This gives $E(a_i \, r_{yi} \, y_i) = \frac{n}{N} p_y \, y_i$

Similarly, $E(a_i \, r_{xi} \, x_i) = \frac{n}{N} p_x \, x_i$.

2. $E(r_{yi} \, r_{yj} \, y_i \, y_j / a_i \, a_j = 1) = p_y^2 \, y_i y_j$ for $i \neq j$.

Result due to assumption 4. This gives

$E(a_i \, a_j \, r_{yi} \, r_{yj} \, y_i \, y_j) = \frac{n(n-1)}{N(N-1)} p_y^2 \, y_i \, y_j$ for $i \neq j$.

The same holds for variable x.

3. $E(r_{yi}\, r_{xi}\, y_i\, x_i/a_i = 1) = p\, y_i\, x_i$, since $y_i\, x_i = 0$ when $i \neq 0$.

This gives $E(a_i\, r_{yi}\, r_{xi}\, y_i\, x_i) = \frac{n}{N}\, p\, y_i\, x_i$.

4. $E(r_{yi}\, r_{xj}\, y_i\, x_j/a_i\, a_j = 1) = p_y\, p_x\, y_i\, x_j$ for $i \neq j$.

Result due to assumption 4. This gives

$E(a_i\, a_j\, r_{yi}\, r_{xj}\, y_i\, x_j) = \frac{n(n-1)}{N(N-1)}\, p_y\, p_x\, y_i\, x_j$ for $i \neq j$.

Hence an unbiased estimator of Y is

$$\hat{Y} = \frac{N}{np_y} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i \; . \qquad \ldots \text{(I)}$$

The proof is directly obtained from result 1.

5. An unbiased estimator of $\sum_{i=1}^{N} y_i^2$ is

$$\frac{N}{np_y} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 \; .$$

6. An unbiased estimator of $\sum\sum_{i \neq j} y_i\, y_j$ is

$$\frac{(N(N-1)}{n(n-1)p_y^2} \sum\sum_{i \neq j} a_i\, a_j\, r_{yi}\, r_{yj}\, y_i\, y_j \; .$$

The proof is directly obtained from result 2.

Hence an unbiased estimator of the variance of $\hat{Y}$ is

$$v(\hat{Y}) = \frac{N[N-1-(n-1)p_y]}{n(n-1)\, p_y^2} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 - \frac{N-n}{N(n-1)}\, \hat{Y}^2 \qquad \ldots \text{(II)}$$

proof: $V(\hat{Y}) = E(\hat{Y}^2) - [E(\hat{Y})]^2 = E(\hat{Y}^2) - Y^2$.

$$Y^2 = \sum_{i=1}^{N} y_i^2 + \sum\sum_{i \neq j} y_i\, y_j, \text{ which are estimated by 5 and 6.}$$

Also, an unbiased estimator of $E\, \hat{Y}^2$ is $\hat{Y}^2$. Combining the three estimators gives

$$v(\hat{Y}) = \hat{Y}^2 - \frac{N}{np_y} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 - \frac{(N(N-1)}{n(n-1)p_y^2} \sum\sum_{i \neq j} a_i\, a_j\, r_{yi}\, r_{yj}\, y_i\, y_j$$

$$= \hat{Y}^2 - \left[\frac{N}{np_y} - \frac{N(N-1)}{n(n-1)p_y^2}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 - \frac{N(N-1)}{n(n-1)p_y^2} \left[\sum_{i=1}^{N} a_i\, r_{yi}\, y_i\right]^2$$

$$= \frac{N[N-1-(n-1)p_y]}{n(n-1)p_y^2} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 + \left[1 - \frac{n(N-1)}{N(n-1)}\right] \hat{Y}^2$$

which gives formula (II).

The above results also hold for variable x when r_{xi}, p_x and x_i are substituted for r_{yi}, p_y and y_i.

7. An unbiased estimator of $\sum_{i=1}^{N} y_i\, x_i$ is

$$\frac{N}{np} \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i \,.$$

The proof is directly obtained from 3.

8. An unbiased estimator of $\sum\sum_{i\neq j} y_i\, x_j$ is

$$\frac{N(N-1)}{n(n-1)p_y p_x} \sum\sum_{i\neq j} a_i\, a_j\, r_{yi}\, r_{xj}\, y_i\, x_j \,.$$

The proof is directly obtained from result 4.

Hence an unbiased estimator of the covariance between $\hat{Y}$ and $\hat{X}$ is

$$\mathrm{cov}(\hat{Y},\hat{X}) = \frac{N}{n}\left[\frac{N-1}{(n-1)\,p_y p_x} - \frac{1}{p}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i - \frac{N-n}{N(n-1)}\, \hat{Y}\hat{X}. \qquad \ldots \text{(III)}$$

proof: $\mathrm{Cov}(\hat{Y},\hat{X}) = E(\hat{Y}\hat{X}) - E(\hat{Y})\,E(\hat{X}) = E(\hat{Y}\hat{X}) - YX.$

$YX = \sum_{i=1}^{N} y_i\, x_i + \sum\sum_{i\neq j} y_i\, x_j$, which are estimated by 7 and 8.

Also, an unbiased estimator of $E(\hat{Y}\hat{X})$ is $\hat{Y}\hat{X}$. This gives

$$\mathrm{cov}(\hat{Y},\hat{X}) = \hat{Y}\hat{X} - \frac{N}{np} \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i -$$

$$\frac{N(N-1)}{n(n-1)p_y p_x} \sum\sum_{i\neq j} a_i\, a_j\, r_{yi}\, r_{xj}\, y_i\, x_j$$

$$= \hat{\hat{Y}}\hat{\hat{X}} - \left[\frac{N}{np} - \frac{N(N-1)}{n(n-1)p_y p_x}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i -$$

$$\frac{N(N-1)}{n(n-1)p_y p_x} \left[\sum_{i=1}^{N} a_i\, r_{yi}\, y_i\right]\left[\sum_{i=1}^{N} a_i\, r_{xi}\, x_i\right]$$

$$= \frac{N}{n}\left[\frac{N-1}{(n-1)p_x p_y} - \frac{1}{p}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i +$$

$$\left[1 - \frac{n(N-1)}{N(n-1)}\right] \hat{\hat{Y}}\hat{\hat{X}}$$

which gives formula (III).

Now the subscript h is introduced to denote the stratum number, assuming that there are L strata in the population. Thus the population total for variable y becomes

$$Y = \sum_{h=1}^{L} Y_h ,$$

and is estimated by

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h , \qquad \ldots \text{(IV)}$$

with each $\hat{Y}_h$ calculated as in formula (I).

The variance of $\hat{Y}$ is estimated by

$$v(\hat{Y}) = \sum_{h=1}^{L} v(\hat{Y}_h) \qquad \ldots \text{(V)}$$

with each $v(\hat{Y}_h)$ calculated as in Formula II.

To obtain estimators of ratios between two totals Y and X, standard statistical estimators and variances are used. These are found in most sampling theory textbooks.
(For example, see chapter 10 of Sampling Theory and Methods by M.N. Murthy (1967).

An estimator of the ratio Y/X is $\hat{Y}/\hat{X}$, with $\hat{X}$ defined the same way as $\hat{Y}$. ... (VI)

This estimator is slighly biased, the bias becoming negligible for large sample sizes. An estimator of the variance of $\hat{Y}/\hat{X}$ is

$$v(\hat{Y}/\hat{X}) = \frac{1}{\hat{X}^2}\left[v(\hat{Y}) - 2\frac{\hat{Y}}{\hat{X}}\,\mathrm{cov}(\hat{Y},\hat{X}) + \left(\frac{\hat{Y}}{\hat{X}}\right)^2 v(\hat{X})\right], \quad \ldots \text{(VII)}$$

with $v(\hat{X}) = \sum_{h=1}^{L} v(\hat{X}_h)$ and $\mathrm{cov}(\hat{Y},\hat{X}) = \sum_{h=1}^{L} \mathrm{cov}(\hat{Y}_h, \hat{X}_h)$

calculated as in formulae II and III.

Finally, since p_y, p_x and p are not known values they are estimated from the sample. The estimates are (within each stratum):

$p_y = \sum_{i\varepsilon I} a_i r_{yi} / \sum_{i\varepsilon I} a_i$, i.e. the proportion responding to y among the sample farms that are in business.

$p_x = \sum_{i\varepsilon I} a_i r_{xi} / \sum_{i\varepsilon I} a_i$, i.e. the proportion responding to x among the sample farms that are in business.

and $p = \sum_{i\varepsilon I} a_i r_{yi} r_{xi} / \sum_{i\varepsilon I} a_i$, i.e. the proportion responding to x and y among the sample farms that are in business.

B-108F
c.2.

Statistics Canada Statistique Canada

Ottawa

PLAN DE L'ENQUÊTE SUR LA VOLAILLE ET LES OEUFS, 1982

K. Smith-Doiron
J.-L. Tambay
DMEIA

Table des matières

1. Introduction 1

2. Base de sondage 1

3. Plan de l'échantillon 1

4. Chevauchement avec l'ENF 2

5. Méthodes d'estimation 2

Annexe I - Caractéristiques du plan de l'échantillon ... 4

Annexe II - Formules d'estimation 6

Plan de l'enquête sur la volaille et les oeufs, 1982

1. Introduction

L'enquête sur la volaille et les oeufs était à l'origine une enquête postale mensuelle dont le taux de réponse était faible. L'enquête remaniée (depuis octobre 1982) a lieu trimestriellement par courrier avec suivi téléphonique des cas de non-réponse (sauf pour les fermes qui comptent moins de 18 pondeuses). Il existe deux types de fermes dans l'industrie de la volaille : les fermes enregistrées et les fermes non enregistrées. Les troupeaux enregistrés sont ceux dont le nombre de pondeuses est égal ou supérieur au nombre minimum nécessaire pour être enregistré dans chaque province : 500 à Terre-Neuve, en Nouvelle-Écosse, en Ontario, au Manitoba et en Colombie-Britannique; 300 à l'Île-du-Prince-Édouard, en Saskatchewan et en Alberta; 250 au Québec et 200 au Nouveau-Brunswick. Aucune enquête ne sera menée sur les troupeaux enregistrés des provinces de l'Atlantique à cause du petit nombre de producteurs visés dans ces provinces et parce que les associations provinciales possèdent des renseignements à jour dans ce domaine.

On renouvellera 25 % de l'échantillon à chaque mois d'octobre, sauf dans le cas des strates de dénombrement complet. Le plan et la taille de l'échantillon devraient produire une valeur prévue de 5 % pour le CV des estimations du nombre de pondeuses par province et par troupeau enregistré ou non enregistré. Il se peut que le CV des autres variables ne corresponde pas au même intervalle.

2. Base de sondage

Le recensement de l'agriculture de 1981 a servi de base. Les réserves indiennes, les institutions et les fermes n'ayant aucune pondeuse ont été supprimées de l'échantillon. Dans les provinces de l'Atlantique, on a aussi éliminé les troupeaux enregistrés.

3. Plan de l'échantillon

Troupeaux non enregistrés

Dans chaque province, les troupeaux non enregistrés ont été stratifiés à l'aide des classifications utilisées dans les publications du recensement. La taille de l'échantillon a été répartie selon la méthode de répartition optimale. Dans un cas (strate 4 à Terre-Neuve), la répartition optimale exigeait un dénombrement complet. Les fermes de cette strate sont désignées des fermes spécifiées.

Troupeaux enregistrés

Les troupeaux enregistrés ont été divisés en trois strates à l'intérieur de chaque province. Les limites des strates ont été précisées d'une manière optimale au moyen des méthodes de répartition optimale. Encore une fois, cela donne comme résultat que certaines strates sont complètement dénombrées (comme la strate 6 au Québec, la strate 7 en Ontario et la strate 7 en Saskatchewan).

La taille de l'échantillon (sauf dans les strates de dénombrement complet) est un multiple de 4 afin de rendre possible le renouvellement de 25 % de l'échantillon. Il n'y a pas de renouvellement dans les strates de dénombrement complet. Un échantillon aléatoire simple a été prélevé dans chaque strate. La définition ainsi que la taille de la population et de l'échantillon de chaque strate sont présentées au tableau I de l'annexe I.

4. Chevauchement avec l'enquête nationale sur les fermes

L'échantillon de l'enquête sur la volaille sera apparié avec les unités spécifiées et les éléments sélectionnés de l'échantillon de l'ENF de 1983. Toutes les fermes avicoles qui figurent dans les deux échantillons seront repérées et leur questionnaire de juillet sera ajouté aux documents de l'ENF. Ainsi, les interviewers rempliront les questionnaires sur la volaille en même temps que ceux de l'ENF. Au cours des années subséquentes, il sera de nouveau possible de trouver les fermes qui participent à l'ENF et à l'enquête sur la volaille et d'inclure les deux questionnaires dans les documents des interviewers de l'ENF.

5. Méthodes d'estimation

Les techniques d'estimation et les formules utilisées pour le calcul des estimations et de leur variation sont résumées à l'annexe II.

Les estimateurs prennent en compte le plan de l'échantillon de l'enquête, de même que le taux prévu de non-réponse. Les estimateurs relatifs au plan d'échantillonnage (c'est-à-dire un échantillon aléatoire simple à l'intérieur de chaque strate) ont été intégrés dans un modèle de non-réponse afin de produire les estimations de l'enquête. Une brève description du modèle de non-réponse est donnée ci-dessous.

Le modèle suppose que, pour chaque ferme, et une fois qu'on est entré en contact avec le répondant, il existe une certaine probabilité d'obtenir une réponse ou une série de réponses aux divers éléments du questionnaire. Cette probabilité n'est peut-être pas la même pour toutes les parties du questionnaire ou pour toutes les fermes visées par l'enquête, mais elle ne variera pas au cours de la période d'inclusion d'une ferme dans l'échantillon. Le modèle de non-réponse impose aussi certaines conditions concernant les valeurs possibles de la probabilité d'une réponse.

La première condition est qu'on obtient une réponse de toutes les fermes qui ont cessé leur exploitation commerciale et dont on a contacté le répondant, les réponses étant toutes zéro. Ce raisonnement repose sur la supposition que toutes les fermes dans l'échantillon qui ont cessé leur exploitation commerciale seront bien repérées à l'étape de la collecte des données; la probabilité d'une réponse est donc toujours égale à un dans ce cas. La deuxième condition est que toutes les fermes d'une strate, à l'exception des fermes qui ont cessé leur exploitation commerciale, ont la même probabilité de répondre à certains éléments ou groupes d'éléments du questionnaire. Essentiellement, on suppose que les fermes qui composent une strate possèdent certaines caractéristiques communes qui produisent un comportement semblable d'une ferme à l'autre dans l'échantillon. Enfin, il est supposé que le genre de réponse d'une ferme à l'égard d'une question ou d'un groupe de questions n'influe nullement sur le type de déclaration d'une autre ferme dans l'échantillon au sujet des mêmes questions. Ces hypothèses devraient normalement être vraies dans la plupart des enquêtes.

Pendant l'élaboration des formules des estimations de l'enquête, les probabilités d'une réponse ont été considérées comme des constantes connues. En réalité, il faut estimer ces probabilités à partir des données sur les répondants de l'enquête. Toutefois, cette simplification n'est pas censée entraîner de grandes modifications des chiffres de l'enquête.

ANNEXE I

TABLEAU I

PLAN D'ÉCHANTILLONNAGE DE L'ENQUÊTE SUR LA VOLAILLE ET LES OEUFS, 1982

PROVINCE	GENRE DE FERME[1]	STRATE	NOMBRE DE PONDEUSES	TAILLE DE LA POPULATION	TAILLE DE L'ÉCHANTILLON
Terre-Neuve	NE	1	1-17	58	4
		2	18-77	43	12
		3	78-249	7	4
		4	250-499	6	6
		TOTAL		114	26
Île-du-Prince-Édouard	NE	1	1-17	256	8
		2	18-77	217	20
		3	78-299	63	24
		TOTAL		536	52
Nouvelle-Écosse	NE	1	1-77	512	12
		2	18-77	450	24
		3	78-249	56	12
		4	250-499	18	8
		TOTAL		1,036	56
Nouveau-Brunswick	NE	1	1-17	326	8
		2	18-77	426	28
		3	78-199	96	16
		TOTAL		848	52
Québec	NE	1	1-17	2,905	12
		2	18-77	3,461	40
		3	78-249	517	20
		total		6,883	72
	E	4	250-11,569	269	32
		5	11,570-52,329	89	28
		6	52,330 et plus	9	9
		total		367	69
		TOTAL		7,250	141
Ontario	NE	1	1-17	6,559	8
		2	18-77	6,767	20
		3	78-249	1,503	12
		4	250-499	652	12
		total		15,481	52
	E	5	500-8,254	703	36
		6	8,255-36,658	281	36
		7	36,659 et plus	28	28
		total		1,012	100
		TOTAL		16,493	152

[1] E = troupeaux enregistrés
NE = troupeaux non enregistrés

PROVINCE	GENRE DE FERME[1]	STRATE	NOMBRE DE PONDEUSES	TAILLE DE LA POPULATION	TAILLE DE L'ÉCHANTILLON
Manitoba	NE	1	1-17	1,337	4
		2	18-77	2,445	28
		3	78-249	347	12
		4	250-499	121	8
		total		4,250	52
	E	5	500-6,151	163	16
		6	6,152-16,977	88	12
		7	16,978 et plus	31	24
		total		282	52
		TOTAL		4,532	104
Saskat-chewan	NE	1	1-17	4,286	8
		2	18-77	8,507	48
		3	78-299	715	16
		total		13,508	72
	E	4	300-3,558	92	12
		5	3,559-16,595	45	16
		6	16,596-114,999	9	4
		7	115,000 et plus	1	1
		total		147	33
		TOTAL		13,655	105
Alberta	NE	1	1-17	3,575	8
		2	18-77	8,146	40
		3	78-299	1,159	24
		total		12,880	72
	E	4	300-3,701	260	20
		5	3,702-13,564	140	28
		6	13,565 et plus	28	16
		total		428	64
		TOTAL		13,308	136
Colombie-Britan-nique	NE	1	1-17	2,888	12
		2	18-77	3,369	36
		3	78-249	265	8
		4	250-499	69	8
		total		6,591	64
	E	5	500-10,212	99	12
		6	10,213-21,862	110	12
		7	21,863 et plus	16	8
		total		225	32
		TOTAL		6,816	96
Canada	NE+E	TOTAL		64,588	920

[1] E = troupeaux enregistrés
NE = troupeaux non enregistrés

ANNEXE II : FORMULES D'ESTIMATION

On décrit ici le calcul des estimations et de leur variance à l'intérieur du plan d'échantillonnage et du modèle de non-réponse utilisés. La section A de cette annexe définit la notation, la section B présente les hypothèses du modèle et la section C résume les résultats théoriques qui conduisent aux formules IV à VII.

Pour simplifier la notation, on supposera que l'échantillon contient seulement une strate, sauf indication contraire.

A. Notation:

Y_i - valeur de la variable y pour la $i^{ième}$ unité de la population ($i = 1, 2, \ldots, N$)

x_i - valeur de la variable x pour la $i^{ième}$ unité de la population ($i = 1, 2, \ldots, N$)

N - taille de la population

n - taille de l'échantillon

a_i - variable aléatoire utilisée pour indiquer si l'unité i figure dans l'échantillon

$$a_i = \begin{cases} 1 & \text{si l'unité i est dans l'échantillon} \\ 0 & \text{autrement} \end{cases}$$

NOTE: $\sum_{i=1}^{N} a_i = n$

r_{yi} - variable aléatoire de réponse qui indique si l'unité i de l'échantillon a répondu à la question sur la variable y

$$r_{yi} = \begin{cases} 1 & \text{si l'unité i a répondu à la variable y} \\ 0 & \text{autrement} \end{cases}$$

r_{xi} - variable aléatoire de réponse qui indique si l'unité i de l'échantillon a répondu à la question sur la variable x

$$r_{xi} = \begin{cases} 1 & \text{si l'unité i a répondu à la variable x} \\ 0 & \text{autrement} \end{cases}$$

Y - total de la variable y pour l'ensemble de la population

$$Y = \sum_{i=1}^{N} Y_i$$

X - total de la variable x pour l'ensemble de la population

$$X = \sum_{i=1}^{N} x_i$$

B. Hypothèses:

1. L'enquête est basée sur un échantillon aléatoire simple sans remise. On obtient les résultats habituels ci-dessous:

$$Pr(a_i=1) = \frac{n}{N}$$

$$Pr(a_i a_j=1) = \frac{n(n-1)}{N(N-1)} \text{ for } i \neq j.$$

2. Chaque unité de la population appartient à un groupe parmi deux, lesquels sont représentés par les symboles I et O qui dénotent respectivement les fermes actives et les fermes qui ont cessé leur exploitation. La valeur des variables y et x est toujours nulle pour les unités du groupe O et ces unités répondent toujours aux questions relatives aux variables y et x. Bref, pour i O, $Y_i = x_i = 0$ et

$$r_{yi} = r_{xi} = 1.$$

3. Pour les fermes du groupe I, quand $a_i = 1$

$$Pr(r_{yi} = 1 \mid i \in I) = p_y,$$

$$Pr(r_{xi} = 1 \mid i \in I) = p_x, \text{ et}$$

$$Pr(r_{yi} = 1, r_{xi} = 1 \mid i \in I) = p.$$

4. Pour une valeur donnée de a_i et de a_j, les variables aléatoires de réponse r_{yi} et r_{xi} sont indépendantes des variables aléatoires de réponse r_{yj} et r_{xj} pour $i \neq j$.

C. Résultats :

1. $E(r_{yi}y_i/a_i=1) = p_y y_i$, puisque $y_i = 0$ quand $i \varepsilon 0$.

   Ainsi, $E(a_i r_{yi} y_i) = \frac{n}{N} p_y y_i$

   De même, $E(a_i r_{xi} x_i) = \frac{n}{N} p_x x_i$.

2. $E(r_{yi}\, r_{yj}\, y_i\, y_j / a_i\, a_j = 1) = p_y^2\, y_i y_j$ pour $i \neq j$.
   Ce résultat découle de l'hypothèse 4. Il s'ensuit que
   $E(a_i\, a_j\, r_{yi}\, r_{yj}\, y_i\, y_j) = \frac{n(n-1)}{N(N-1)}\, p_y^2\, y_i y_j$ pour $i \neq j$.
   Il en va de même pour la variance x.

3. $E(r_{yi}\, r_{xi}\, y_i\, x_i / a_i = 1) = p\, y_i\, x_i$, puisque $y_i\, x_i = 0$ quand $i \varepsilon 0$.

   Donc, $E(a_i\, r_{yi}\, r_{xi}\, y_i\, x_i) = \frac{n}{N} p\, y_i\, x_i$.

4. $E(r_{yi}\, r_{xj}\, y_i\, x_j / a_i\, a_j = 1) = P_y\, p_x\, y_i\, x_j$ pour $i \neq j$.
   Ce résultat découle de l'hypothèse 4. Il s'ensuit que
   $E(a_i\, a_j\, r_{yi}\, r_{xj}\, y_i\, x_j) = \frac{n(n-1)}{N(N-1)}\, p_y\, p_x\, y_i\, x_j$ pour $i \neq j$.

Ainsi, la formule suivante est celle d'un estimateur sans biais de Y:

$$\hat{Y} = \frac{N}{np_y} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i \; . \qquad \ldots \text{ (I)}$$

Le résultat 1 conduit à une preuve directe de cette conclusion.

5. L'équation suivante produit une estimation sans biais de $\sum_{i=1}^{N} y_i^2$:

$$\frac{N}{np_y} \sum_{i=1}^{N} a_i\, r_{yi}\, y_i^2 \; .$$

6. L'équation suivante produit une estimation sans biais de $\sum\sum_{i\neq j} y_i y_j$:

$$\frac{(N(N-1))}{n(n-1)p_y^2} \sum\sum_{i\neq j} a_i a_j r_{yi} r_{yj} y_i y_j .$$

Le résultat 2 conduit à une preuve directe de cette conclusion.

Ainsi, la formule suivante est celle d'un estimateur sans biais de la variance de Y :

$$\hat{V}(Y) = \frac{N[N-1-(n-1)p_y]}{n(n-1)p_y^2} \sum_{i=1}^{N} a_i r_{yi} y_i^2 - \frac{N-n}{N(n-1)} \hat{Y}^2 \quad \ldots \text{(II)}$$

preuve : $V(\hat{Y}) = E(\hat{Y}^2) - [E(\hat{Y})]^2 = E(\hat{Y}^2) - Y^2$.

$Y^2 = \sum_{i=1}^{N} Y_i^2 + \sum\sum_{i=1} Y_i Y_j$; les deux termes de cette équation sont estimés au moyen des formules présentées aux résultats 5 et 6. En outre, $\hat{Y}^2$ est un estimateur sans biais de $E(\hat{Y}^2)$. Si l'on réunit ces trois estimateurs, on obtient l'équation suivante :

$$\hat{V}(Y) = \hat{Y}^2 - \frac{N}{np_y} \sum_{i=1}^{N} a_i r_{yi} y_i^2 - \frac{(N(N-1))}{n(n-1)p_y^2} \sum\sum_{i\neq j} a_i a_j r_{yi} r_{yj} y_i y_j$$

$$= \hat{Y}^2 - \left[\frac{N}{np_y} - \frac{N(N-1)}{n(n-1)p_y^2}\right] \sum_{i=1}^{N} a_i r_{yi} y_i^2 - \frac{N(N-1)}{n(n-1)p_y^2} \left[\sum_{i=1}^{N} a_i r_{yi} y_i\right]^2$$

$$= \frac{N[N-1-(n-1)p_y]}{n(n-1)p_y^2} \sum_{i=1}^{N} a_i r_{yi} y_i^2 + \left[1 - \frac{n(N-1)}{N(n-1)}\right] \hat{Y}^2$$

Ce qui correspond à la formule (II).

Les résultats décrits plus haut s'appliquent également à la variable x si l'on remplace r_{yi} , p_y et y_i par r_{xi} , p_x et x_i.

7. L'équation suivante produit une estimation sans biais de $\sum_{i=1}^{N} y_i x_i$:

$$\frac{N}{np} \sum_{i=1}^{N} a_i r_{yi} r_{xi} y_i x_i .$$

Le résultat 3 conduit à une preuve directe de cette conclusion.

8. L'équation suivante produit une estimation sans biais de $\sum\sum_{i \neq j} y_i x_j$:

$$\frac{N(N-1)}{n(n-1)p_y p_x} \sum\sum_{i \neq j} a_i a_j r_{yi} r_{xj} y_i x_j .$$

Le résultat 4 conduit à une preuve directe de cette conclusion.

Ainsi, la formule suivante est celle d'un estimateur sans biais de la covariance entre $\hat{\bar{y}}$ et $\hat{\bar{x}}$:

$$\text{cov}(\hat{\bar{y}},\hat{\bar{x}}) = \frac{N}{n}\left[\frac{N-1}{(n-1)\ p_y p_x} - \frac{1}{p}\right] \sum_{i=1}^{N} a_i r_{yi} r_{xi} y_i x_i - \frac{N-n}{N(n-1)}\ \hat{\bar{y}}\hat{\bar{x}}. \qquad \ldots(III)$$

preuve: $\text{Cov}(\hat{\bar{y}},\hat{\bar{x}}) = E(\hat{\bar{y}}\hat{\bar{x}}) - E(\hat{\bar{y}})\ E(\hat{\bar{x}}) = E(\hat{\bar{y}}\hat{\bar{x}}) - \bar{y}\bar{x}.$

$$\bar{y}\bar{x} = \sum_{i=1}^{N} y_i x_i + \sum\sum_{i \neq j} y_i x_j,$$ les deux termes de cette équation sont estimés au moyen des formules présentées aux résultats 7 et 8. En outre, $\hat{\bar{y}}\hat{\bar{x}}$ est un estimateur sans biais de $E(\hat{\bar{y}}\hat{\bar{x}})$. Si l'on réunit ces trois estimateurs, on obtient l'équation suivante:

$$\text{cov}(\hat{\bar{y}},\hat{\bar{x}}) = \hat{\bar{y}}\hat{\bar{x}} - \frac{N}{np} \sum_{i=1}^{N} a_i r_{yi} r_{xi} y_i x_i -$$

$$\frac{N(N-1)}{n(n-1)p_y p_x} \sum\sum_{i \neq j} a_i a_j r_{yi} r_{xj} y_i x_j$$

$$= \hat{y}\hat{x} - \left[\frac{N}{np} - \frac{N(N-1)}{n(n-1)p_y p_x}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i -$$

$$\frac{N(N-1)}{n(n-1)p_y p_x} \left[\sum_{i=1}^{N} a_i\, r_{yi}\, y_i\right] \left[\sum_{i=1}^{N} a_i\, r_{xi}\, x_i\right]$$

$$= \frac{N}{n}\left[\frac{N-1}{(n-1)p_x p_y} - \frac{1}{p}\right] \sum_{i=1}^{N} a_i\, r_{yi}\, r_{xi}\, y_i\, x_i +$$

$$\left[1 - \frac{n(N-1)}{N(n-1)}\right] \hat{y}\hat{x}$$

Ce qui correspond à la formule (III).

Introduisons maintenant l'indice h pour représenter le numéro de strate, et supposons que la population compte L strates. La valeur totale de la variable y pour l'ensemble de la population s'exprime alors de la manière suivante:

$$y = \sum_{h=1}^{L} y_h,$$

et l'estimation de ce total est calculée au moyen de la formule:

$$\hat{y} = \sum_{h=1}^{L} \hat{y}_h, \qquad \ldots(IV)$$

où chaque $\hat{y}_h$ est calculé à partir de la formule (I).

On peut estimer la variance de $\hat{y}$ à l'aide de l'équation :

$$v(\hat{y}) = \sum_{h=1}^{L} v(\hat{y}_h) \qquad \ldots(V)$$

où chaque $v(\hat{y}_h)$ est calculé à partir de la formule (II).

Pour obtenir des estimations du rapport entre deux totaux y et x, on utilise les formules statistiques habituelles pour les estimateurs et les variances. Ces formules figurent dans la plupart des manuels sur la théorie de l'échantillonnage. (Voir, par exemple, le chapitre 10 de Sampling Theory and Methods de M.N. Murthy (1967)).

Un estimateur du ratio y/x est $\hat{x}/\hat{y}$, où $\hat{x}$ est défini de façon semblable à $\hat{y}$. ...(VI)

Cet estimateur est légèrement biaisé et le biais devient négligeable dans les échantillons de grande taille. Un estimateur de la variance $\hat{\bar{y}}/\hat{\bar{x}}$ est calculé par la formule suivante:

$$v(\hat{\bar{y}}/\hat{\bar{x}}) = \frac{1}{\hat{\bar{x}}^2}\left[v(\hat{\bar{y}}) - 2\,\frac{\hat{\bar{y}}}{\hat{\bar{x}}}\,\mathrm{cov}(\hat{\bar{y}},\hat{\bar{x}}) + \left(\frac{\hat{\bar{y}}}{\hat{\bar{x}}}\right)^2 v(\hat{\bar{x}})\right], \quad \ldots \text{(VII)}$$

où $v(\hat{\bar{x}}) = \sum_{h=1}^{L} v(\hat{\bar{x}}_h)$ et $\mathrm{cov}(\hat{\bar{y}},\hat{\bar{x}}) = \sum_{h=1}^{L} \mathrm{cov}(\hat{\bar{y}}_h, \hat{\bar{x}}_h)$

calculés au moyen des équations (II) et (III).

Enfin, comme les valeurs de P_y, P_x et p sont inconnues, il faut les estimer de l'échantillon. On obtient donc les estimations suivantes (à l'intérieur de chaque strate) :

$P_y = \sum_{i\in I} a_i r_{yi} / \sum_{i\in I} a_i$, ce qui mesure la proportion des fermes actives dans l'échantillon qui ont répondu à la question sur la variable y.

$P_x = \sum_{i\in I} a_i r_{xi} / \sum_{i\in I} a_i$, ce qui mesure la proportion des fermes actives dans l'échantillon qui ont répondu à la question sur la variable x.

et $p = \sum_{i\in I} a_i r_{yi} r_{xi} / \sum_{i\in I} a_i$, ce qui mesure la proportion des fermes actives dans l'échantillon qui ont répondu aux questions sur les variables x et y.